



Transcriptomics resources of human tissues and organs

Uhlén, Mathias; Hallström, Björn M.; Lindskog, Cecilia; Mardinoglu, Adil; Pontén, Fredrik; Nielsen, Jens

Published in:
Molecular Systems Biology

Link to article, DOI:
[10.15252/msb.20155865](https://doi.org/10.15252/msb.20155865)

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Uhlén, M., Hallström, B. M., Lindskog, C., Mardinoglu, A., Pontén, F., & Nielsen, J. (2016). Transcriptomics resources of human tissues and organs. *Molecular Systems Biology*, 12(4), [862].
<https://doi.org/10.15252/msb.20155865>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Transcriptomics resources of human tissues and organs

Mathias Uhlen^{1,2,3,*}, Björn M Hallström^{1,2}, Cecilia Lindskog⁴, Adil Mardinoglu^{1,5}, Fredrik Pontén⁴ & Jens Nielsen^{1,3,5}

Abstract

Quantifying the differential expression of genes in various human organs, tissues, and cell types is vital to understand human physiology and disease. Recently, several large-scale transcriptomics studies have analyzed the expression of protein-coding genes across tissues. These datasets provide a framework for defining the molecular constituents of the human body as well as for generating comprehensive lists of proteins expressed across tissues or in a tissue-restricted manner. Here, we review publicly available human transcriptome resources and discuss body-wide data from independent genome-wide transcriptome analyses of different tissues. Gene expression measurements from these independent datasets, generated using samples from fresh frozen surgical specimens and postmortem tissues, are consistent. Overall, the different genome-wide analyses support a distribution in which many proteins are found in all tissues and relatively few in a tissue-restricted manner. Moreover, we discuss the applications of publicly available omics data for building genome-scale metabolic models, used for analyzing cell and tissue functions both in physiological and in disease contexts.

Keywords genome-scale metabolic models; proteomics; transcriptomics

DOI 10.15252/msb.20155865 | Received 28 September 2015 | Revised 15

February 2016 | Accepted 17 February 2016

Mol Syst Biol. (2016) 12: 862

Introduction

The global classification of the human proteome with regard to its spatiotemporal expression patterns and its functions represents one of the major challenges for studying human biology and disease (Lamond *et al.*, 2012). Recently, genomic, transcriptomic, and proteomic technologies have been employed to analyze the human proteome on a genome-wide level. Genome annotation efforts, such as Ensembl (Cunningham *et al.*, 2015) and GenCODE consortium (Harrow *et al.*, 2012) in the Encode project (Nilsson

et al., 2015), have identified approximately 20,000 genes coding for proteins, and the UniProt consortium (UniProt, 2015) has manually annotated the majority of these genes. On the transcript level, expression levels of human genes have been monitored to study the effects of diseases, treatments, and developmental stages using microarray-based gene expression profiling (Brawand *et al.*, 2011; Petryszak *et al.*, 2015). Recently, several efforts have been published with the quantitative analysis of RNA levels based on next-generation sequencing in samples representing most of the major organs and tissues in the human body (Fig 1), including the Fantom consortium (Yu *et al.*, 2015), the Human Protein Atlas (HPA) consortium (Uhlen *et al.*, 2015), and the genome-based tissue expression (GTEx) consortium (Keen & Moore, 2015). On the protein level, several large-scale studies based on mass spectrometry analysis have also been published (Kim *et al.*, 2014; Wilhelm *et al.*, 2014), and these studies have been complemented with antibody-based protein profiling using tissue microarrays containing samples representing most major tissues and organs in the human body (Fagerberg *et al.*, 2014; Uhlen *et al.*, 2015). Most of the quantitative data on the expression of protein-coding genes are based on recent transcriptomics studies based on RNA-seq. Here, we review some of the publicly available human transcriptome resources and discuss tissue data from independent research groups.

An interesting aspect of the integration of omics technologies is the sampling that depends on the sensitivity and the resolution of each technology. Consequently, the analysis of tissue samples is normally performed on a mixture of cell types using transcriptomics and mass spectrometry-based proteomics, whereas *in situ* hybridization techniques, successfully applied for mapping the distribution of transcripts in the brain (Hawrylycz *et al.*, 2012), and more qualitative approaches involving antibody-based profiling allow analyzing single cells in their natural environment to reveal the differences in protein expression levels between neighboring cells. Thus, antibody-based protein profiling complements quantitative transcriptomics and proteomics, based on a mixture of cell types, to reach single-cell resolution in the analysis of gene expression in complex tissues. Recently, single-cell technologies have been

¹ Science for Life Laboratory, KTH - Royal Institute of Technology, Stockholm, Sweden

² Department of Proteomics, KTH - Royal Institute of Technology, Stockholm, Sweden

³ Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark

⁴ Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

⁵ Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

*Corresponding author: Tel: +46 8 5537 8325; E-mail: mathias.uhlen@scilifelab.se

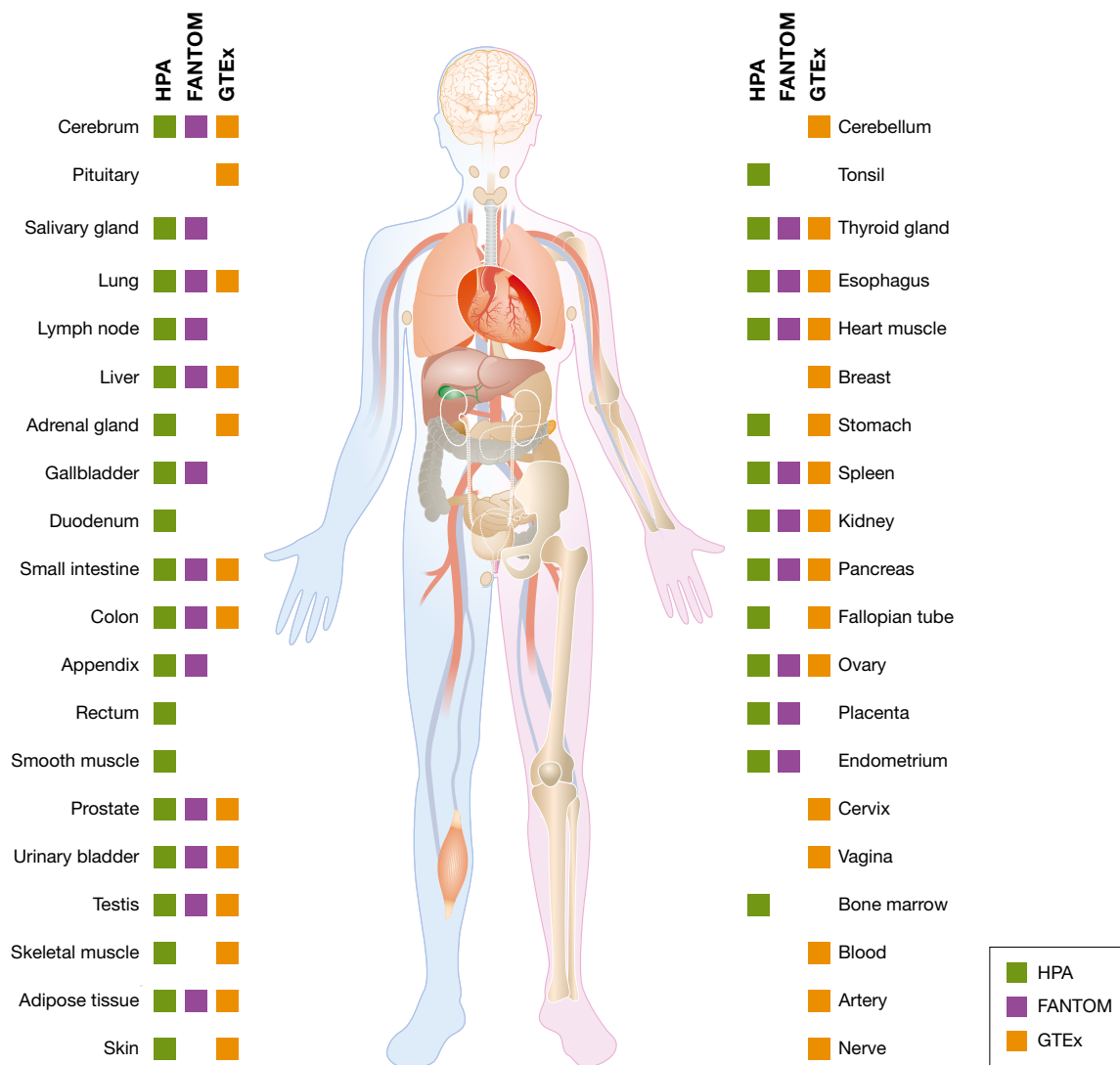


Figure 1. Global transcriptomics analysis of human tissues and organs.

Overview of the tissues and organs analyzed using RNA-seq by the Human Protein Atlas consortium (HPA, green), tissues studied with cap analysis gene expression (CAGE) within the FANTOM consortium (purple), and tissues analyzed using RNA-seq by the genome-based tissue expression consortium (GTEx, orange). Altogether, 22 tissues and organs were studied with both the HPA and FANTOM datasets, while 21 tissues overlapped between the HPA and GTEx datasets.

developed for transcriptomics (Tang *et al*, 2009), but no global analysis across many tissues has yet been published using such methodology. In the near future, the possibility to move to single-cell analyses of tissues for transcriptomics and proteomics will allow analyzing cell-to-cell variability, which is particular interesting, that is, in the context of cancer. Moreover, the development of more quantitative assays for immunohistochemistry using fluorescent probes will be valuable for providing quantitative data for whole-cell modeling (Stadler *et al*, 2013).

The correlation between mRNA levels and the corresponding protein levels is an important issue for the comparability of the different omics-based technologies, and the presence or absence of such correlation on an individual gene/protein level has been debated in the literature for many years (Anderson & Seilhamer, 1997; Tian *et al*, 2004; Gry *et al*, 2009; Maier *et al*, 2009, 2011;

Lundberg *et al*, 2010; Schwanhaussner *et al*, 2011). However, a comparison of steady-state levels of mRNA and proteins in human cell lines using RNA-seq and quantitative triple-SILAC analysis showed good genome-wide correlations when the mRNA and protein levels of an individual gene were compared in three separate cell lines (Lundberg *et al*, 2010). These observations were supported by Kuster and coworkers (Wilhelm *et al*, 2014) comparing mass spectrometry data from different tissues with RNA-seq data obtained from the HPA consortium (Uhlen *et al*, 2015). Again, the steady-state levels for individual genes correlated across several tissues. Overall, these studies suggest that the amount of a given protein in a cell or tissue is, in general, reflected by the corresponding mRNA level, although this gene-/protein-specific ratio may vary greatly between different gene products depending on various factors, mainly translational rates and protein half-lives (Eden *et al*,

2011), and the transcript level for a given gene might therefore be used to predict the corresponding protein level. This hypothesis needs to be confirmed by more in-depth studies using quantitative measurements at both the RNA and protein levels. However, it suggests that next-generation sequencing of the transcriptome is a powerful tool for indirect measurements of protein expression levels, thus forming an attractive link between the field of genomics and proteomics.

Analyses of the human transcriptome by different methods

Several genome-wide transcriptome profiling methods have been used for identifying and quantifying global gene expression patterns, each allowing for a quantitative analysis of RNA transcripts. Whole-body maps based on microarray analyses have been launched, such as BioGPS (Wu *et al*, 2009), and these have recently been followed by several tissue-based data portals based on RNA-seq. Some of these publicly available repositories for transcriptomics data are summarized in Table 1 with a focus on datasets from RNA-seq experiments. The resources include repositories for external data, such as Expression Atlas from the European Bioinformatics Institute (EBI) and Gene Expression Omnibus from the National Center for Biotechnology Information (NCBI), as well as repositories with internally generated transcriptome data, such as the GTEx, the Human Protein Atlas, and the Allen Brain Atlas. In a recent study (Danielsson *et al*, 2015), the concordance of RNA-seq data from four large-scale efforts was compared based on gene expression measurements for ostensibly similar samples (specifically, human brain, heart, and kidney samples). The authors concluded that human tissue RNA-seq expression measurements appear quite consistent, considering that samples cluster by tissue rather than laboratory of origin given simple preprocessing transformations.

An alternative approach to RNA-seq, named cap analysis gene expression (CAGE), has been described by the Fantom consortia (Yu *et al*, 2015) and allows for quantitative measurements of transcripts based on sequencing the 5'-end of capped mRNA molecules. The correlation between RNA-seq and CAGE for transcriptome analysis was recently investigated (Yu *et al*, 2015), and the transcriptome of 22 tissues was analyzed using both methods (Fig 1) based on 79 RNA-seq (HPA) and 27 CAGE (FANTOM) samples. Tissue-to-tissue comparisons showed a high genome-wide correlation between the two datasets (Yu *et al*, 2015). Interestingly, discrepancies between the two datasets can largely be explained by gene model annotation issues or technical artifacts inherent in the respective methodologies. As an example, the HPA data excluded mRNA without poly-adenylation tails and it is therefore not surprising that many histone genes were lacking in the RNA-seq data, but are present in the CAGE data. Conversely, CAGE peaks mapping more than 500 base pairs from the transcriptional start site are lacking in the CAGE dataset, as well as CAGE peaks mapping to two or more locations on the genome, which are removed from the dataset. Thus, the two methods are complementary and it would be attractive to integrate data obtained by these two approaches to refine gene models and to improve the interpretation of gene expression values.

Classification of all human protein-coding genes based on tissue profiling

The different omics-based analyses of the human proteome have allowed the classification of protein-coding genes with regard to tissue-restricted expression. In the analysis performed by the HPA consortium (Fagerberg *et al*, 2014; Uhlen *et al*, 2015), a cutoff of 1 FPKM (Hebenstreit *et al*, 2011) was used to indicate the presence or absence of transcripts for a particular gene in a tissue. Based on this definition, all human protein-coding genes were classified into (i) genes with an elevated expression in one or several tissues, (ii) genes expressed in all analyzed tissues, (iii) genes with mixed expression found in several, but not all tissues, and (iv) genes not detected in any tissues. The elevated genes were further stratified into “tissue enriched”, “group enriched”, or “tissue enhanced”. The term “tissue specific” was avoided as it depends on the definition of cutoff values, and only few genes, including well-known proteins such as insulin, PSA, and troponin, were found to be exclusively expressed in a single tissue type (Uhlen *et al*, 2015). A classification of all protein-coding genes is shown in Table 2 (cutoff of 0.5 FPKM).

The classification of the human protein-coding transcriptome showed that almost half of the genes were detected in all tissues (45%), while 13% showed a mixed expression (Fig 2A). Approximately one-third of the genes showed a tissue elevated expression with 13% of the genes enriched in one of the analyzed tissues. Only 5% of the genes were not detected in any of the analyzed tissues. A further analysis of the number of genes with a tissue elevated expression (Fig 2B) showed that the testis has by far the highest number of tissue-enriched genes followed by the brain (cerebral cortex) and liver.

The recently published RNA-Seq data generated by the GTEx consortium (Bahcall, 2015; GTEx Consortium, 2015; Gibson, 2015) allow for an independent tissue-based classification of the human proteome. The GTEx dataset includes more than 1,600 postmortem samples from mostly overlapping, but in some cases unique, tissues compared to the HPA consortium (Fig 1). For example, the GTEx dataset includes more tissue samples from the brain, blood, and nerves, which are not included in the HPA dataset. As illustrated in Fig 2C, the overall tissue-based classification based on the GTEx dataset and an identical cutoff of 0.5 FPKM is similar to that of the HPA with 45% of the genes expressed in all tissues, 14% showing a mixed expression, 12% being tissue-enriched expression, and 5% of the genes not detected in any of the analyzed tissues. With respect to tissue-elevated genes (Fig 2B and D) the testis is again observed to contain the largest number of tissue-enriched genes, followed by the brain (cerebellum, cortex, and pituitary), skin, and liver. The values for the HPA and the GTEx datasets can be found in Table EV1.

Tissue-enriched genes vs. ubiquitously expressed genes

Tissue-enriched genes identified by the analysis of the HPA data (Uhlen *et al*, 2015) based on the definitions shown in Table 2 can be found for all tissues in the interactive HPA database (www.proteinatlas.org/humanproteome/tissue+specific). A functional Gene Ontology analysis of the tissue-enriched genes in the HPA

Table 1. Data resources for RNA expression data with relevance for human protein-encoding genes.

| Resource | Affiliation | Description | Link (URL) | References |
|------------------------------|--------------------------------------|---|---|---------------------------------|
| Human Protein Atlas | Science for Life Lab (Sweden) | Tissue-based RNA data based on surgically removed tissues (RNA-Seq) | http://www.proteinatlas.org/ | (Uhlen <i>et al</i> , 2015) |
| GTEx | Broad Institute (USA) | Tissue-based RNA data based on postmortem samples (RNA-Seq) | http://gtexportal.org/home/ | (Keen & Moore, 2015) |
| FANTOM | Riken Institute (Japan) | Tissue-based RNA data based on CAGE | http://fantom.gsc.riken.jp/ | (Yu <i>et al</i> , 2015) |
| RNA-Seq Atlas | J. Gutenberg University (Germany) | A reference database for gene expression profiling in normal tissue by next-generation sequencing | http://medicalgenomics.org/rna_seq_atlas | (Krupp <i>et al</i> , 2012) |
| Allen Brain Atlas | Allen Institute (USA) | An anatomically comprehensive atlas of the adult human brain transcriptome | http://human.brain-map.org/ | (Hawrylycz <i>et al</i> , 2012) |
| Evolution of gene expression | University of Lausanne (Switzerland) | The evolution of gene expression levels in mammalian organs | http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30352 | (Brawand <i>et al</i> , 2011) |
| Altiso | MIT (USA) | Alternative isoform regulation in human tissue transcriptomes. | http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12946 | (Wang <i>et al</i> , 2008) |
| Expression Atlas | EBI (UK) | Repository for RNA expression data (both microarray and RNA-Seq) | https://www.ebi.ac.uk/gxa | (Petryszak <i>et al</i> , 2015) |
| ArrayExpress | EBI (UK) | International functional genomics public data repositories | http://www.ebi.ac.uk/arrayexpress/ | (Rustici <i>et al</i> , 2013) |
| Illumina Body Map | Illumina (USA) | RNA-Seq of 16 human individual tissues | http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513/ | (Rustici <i>et al</i> , 2013) |
| Gene Expression Omnibus | NCBI (USA) | Repository for RNA expression data (both microarray and RNA-Seq) | http://www.ncbi.nlm.nih.gov/geo/ | (Barrett <i>et al</i> , 2013) |

Table 2. Classification of all human protein-coding genes based on transcript expression levels in tissues and organs. The columns HPA and GTEx indicate the number of genes identified in the different categories using the datasets (Keen & Moore, 2015; Uhlen *et al*, 2015) from these two consortia.

| Category | Definition | HPA | GTEx |
|------------------|--|--------|--------|
| Tissue enriched | At least fivefold higher mRNA levels (FPKM) in a particular tissue as compared to all other tissues | 2,359 | 2,289 |
| Group enriched | At least fivefold higher mRNA levels in a group of tissues (2–7) | 1,208 | 1,307 |
| Enhanced | At least fivefold higher mRNA levels in a particular tissue as compared to the average levels in all tissues | 3,227 | 3,077 |
| Expressed in all | Detected in all tissues | 8,385 | 8,459 |
| Mixed | Detected in at least two tissues, but not in all, and not part of any of the categories above | 2,484 | 2,537 |
| Not detected | Not present in any of the analyzed tissues (under cutoff) | 1,021 | 1,015 |
| Total | Total number of genes analyzed | 18,684 | 18,684 |
| Total elevated | Total number of tissue-enriched, group-enriched, and tissue-enhanced genes | 6,794 | 6,673 |

dataset has been performed and the results are consistent with the function of each tissue (Uhlen *et al*, 2015). As an example, genes elevated in liver encode secreted plasma and bile proteins, detoxification proteins, and proteins associated with metabolic processes and glycogen storage (Kampf *et al*, 2014), whereas genes elevated in adipose tissue encode proteins involved in lipid metabolic processes (Mardinoglu *et al*, 2014b) and genes elevated in skin encode proteins associated with functions related to the barrier function (squamous cell differentiation and cornification), skin pigmentation, and hair development (Edqvist *et al*, 2015). In order to further validate these lists, we have compared the overlap of tissue-enriched genes identified using the independent HPA and GTEx datasets. The number of tissue-enriched genes in the different tissues and the overlap between the two datasets are shown in Figs 2E and 3A, and Table EV2.

Overall, it is reassuring that there is a significant overlap in the tissue classification of the genes based on the two independent datasets. The fact that similar results are obtained when using fresh frozen tissue (HPA) and postmortem tissue (GTEx) suggests negligible effects of the sampling procedures used by the GTEx consortium on RNA degradation. In the comparison, note that in the HPA dataset, the brain contains only one tissue (cerebral cortex), while the corresponding GTEx dataset is based on three different tissues (cerebellum, cortex, and pituitary). The large discrepancy for skin can be explained by the fact that the sampling of skin in the HPA was based on shave biopsies including mainly epidermis (Uhlen *et al*, 2015), while the GTEx consortium also included the underlying dermis, most likely containing skin adnexal structures such as hair follicles and sweat glands.

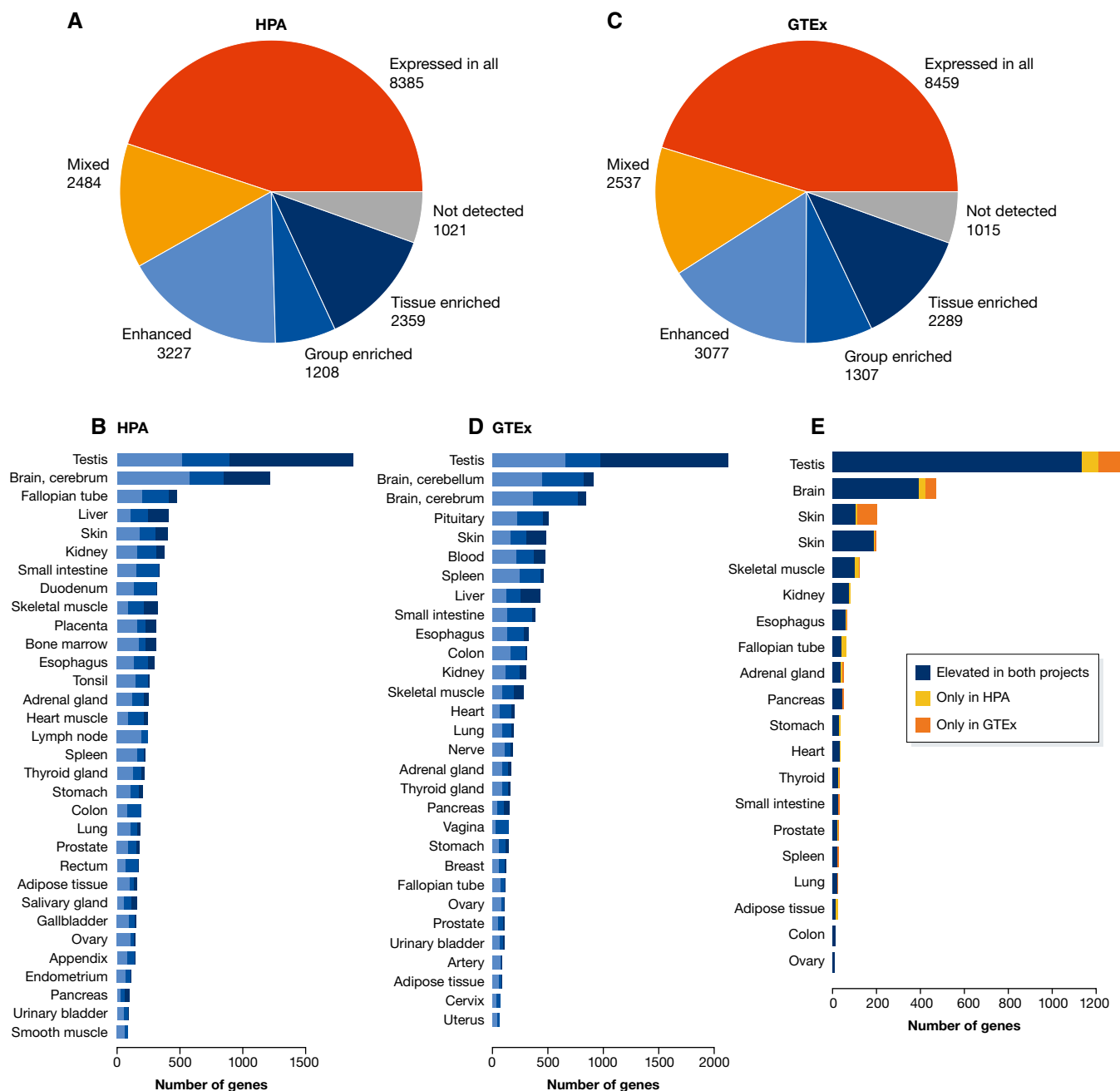


Figure 2. Classification of all protein-coding genes using transcriptomics data.

(A) Pie chart showing the number of genes that fall into each expression specificity category, based on the classifications of HPA (32 tissues, 137 samples) (with a cutoff of 0.5 FPKM). (B) The number of protein-coding genes classified as tissue enriched (dark blue), group enriched (medium blue), and tissue enhanced (light blue) based on the HPA dataset. (C) Pie chart showing the number of genes that fall into each expression specificity category, based on the classifications of GTEx (30 tissues, 2,510 samples) (Keen & Moore, 2015) (with a cutoff of 0.5 FPKM). (D) The number of protein-coding genes classified as above based on GTEx dataset. (E) Barplot showing the overlap of tissue-enriched genes between the two datasets. All genes that are tissue enriched in either dataset are depicted. Genes classified as tissue enriched/group enriched/tissue enhanced in the same tissue in both datasets are shown in blue; genes only enriched in one of the datasets are shown in yellow (only HPA) or orange (only GTEx).

Genes identified as “expressed in all tissues” are expected to be either “housekeeping” genes for which the protein product is needed in every cell, such as proteins involved in transcription, translation, and energy metabolism, or genes expressed in cell types that are present in all tissue types, such as lymphocytes, macrophages, fibroblasts, and endothelial cells. HPA and GTEx

data largely overlap for this category as well, with 7,563 genes identified as “expressed in all” by both resources (Fig 3B). Between 800 and 900 genes were only identified in one of the two datasets, and a more detailed analysis shows that the vast majority of these genes were identified as “mixed” by the other dataset (Fig 3B). This suggests that these genes move between

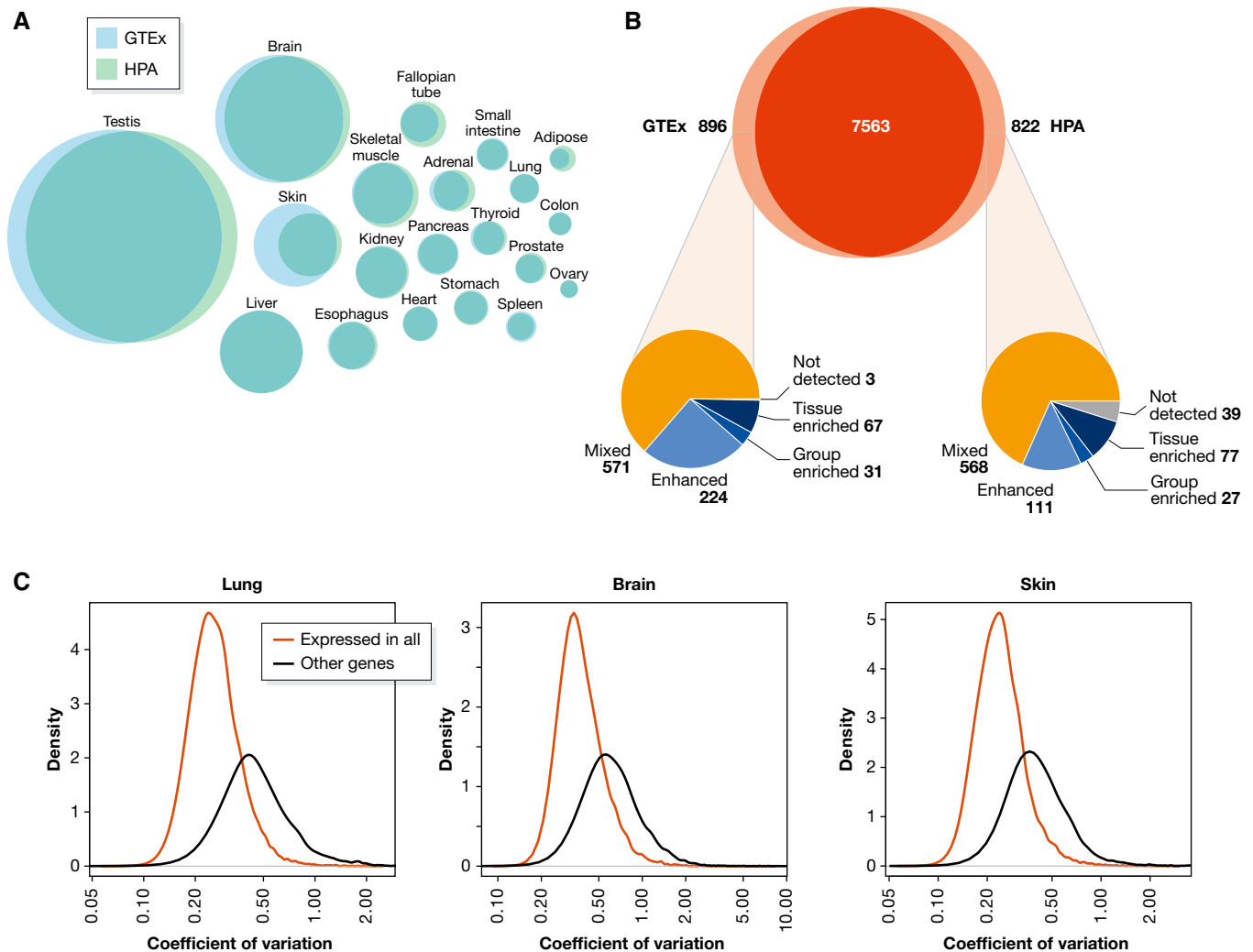


Figure 3. Protein classification and interindividual variations.

(A) Venn diagrams showing the overlap between tissue-elevated genes between the two datasets, HPA in light green and GTEx in light blue. (B) Venn diagram showing the overlap between genes classified as “expressed in all tissues” between the two datasets. The pie charts show the classification of the non-overlapping genes in the dataset where the gene was not detected in all tissues. (C) Comparison of interindividual variation between genes that are annotated as “expressed in all tissues” and all other genes, in lung, brain, and skin (these tissues were selected because they have a large number of biological replicates). The plots illustrate the distribution of the coefficient of variation (CV) within the tissue for all genes in each of the two classes (red: expressed in all, black: other). The CV is shifted toward the lower side in the “expressed in all” category ($P \ll 0.001$), suggesting that genes that are expressed in all tissues have lower variation between individuals.

categories based on the relatively arbitrary FPKM cutoff and indicates that expression in a single tissue below the detection threshold makes a gene move from “expressed in all” to “mixed”.

The variation in protein profiles between individuals

A relevant question arising is the level of interindividual variation in gene expression levels. A comparison of individual variation for “housekeeping” genes (defined as expressed in all analyzed tissues) and genes with a more tissue-restricted expression using GTEx data is presented in Fig 3C for three different tissues (lung, brain, and skin) that are represented by a large number of biological replicates. For all three tissues, the coefficient of variation (CV) in the

“expressed in all” category shifted toward the lower side, suggesting that genes expressed in all tissues seem to vary less between individuals for a particular tissue as compared to genes with a tissue-restricted expression pattern. This illustrates that the proteins found in all tissues are expressed at relatively similar levels across the analyzed tissues, suggesting, as perhaps expected, that these proteins that are involved in “basic functions” are required at similar concentrations in the various tissue types.

Building genome-scale metabolic models for human tissues

High-quality genome-wide proteomics and transcriptomics data can be used for generating and improving context-specific biological

networks including protein–protein interaction (PPI), regulatory, signaling, and metabolic networks (Papin *et al*, 2005; Qian *et al*, 2005; Bossi & Lehner, 2009) in order to gain further insights into the differences in cellular functions across tissues. Genome-scale metabolic models (GEMs) that can be reconstructed directly from proteomics or transcriptomics data are particularly well suited for the analysis of biological functions, since they can be applied to examine the metabolic functions associated with a given cell type. Several studies have recently reported the use of proteomics data to reconstruct GEMs for analyzing metabolic processes across different cell and tissue types in humans (Mardinoglu & Nielsen, 2015; O'Brien *et al*, 2015; Yizhak *et al*, 2015; Bjornson *et al*, 2016) and mice (Mardinoglu *et al*, 2015b). GEMs contain thousands of biochemical reactions and their catalyzing protein-coding genes in a cell/tissue, which generate a complex network of molecular interactions capturing the metabolic functions of this cell/tissue (Fig 4A). This reaction network is converted into a computational model using a stoichiometric (S) matrix and can be applied for the analysis of physiological data collected from both healthy and diseased states (Mardinoglu & Nielsen, 2012; Mardinoglu *et al*, 2013b).

The first human GEMs, Recon1 (Duarte *et al*, 2007) and EHMN (Ma *et al*, 2007), were developed nearly 10 years ago and have now grown to the level where they can be used for predicting the metabolic response of cell/tissue to a given perturbation. These integrative models allowed the identification of new drug target candidates by theoretical analyses (Frezza *et al*, 2011), and many enzymes have already been proposed as drug targets for cancer treatment (Yizhak *et al*, 2015). Recently, more comprehensive generic human GEMs including Recon2 (Thiele *et al*, 2013) and HMR2 (Mardinoglu *et al*, 2014a) were constructed by integrating the components of the first generic human GEMs with manually reconstructed context-specific GEMs. Recon2 covers the content of the HepatoNet, a manually reconstructed GEM for hepatocytes (Gille *et al*, 2010), whereas HMR2 covers the content of both HepatoNet and *iAdipocytes1809*, a manually reconstructed GEM for adipocytes (Mardinoglu *et al*, 2013a). HMR2 also includes the extensive description of lipid metabolism present in human adipocytes and hepatocytes. The number of reactions, metabolites, and genes incorporated in each model as well as the conceptual evolution of global reconstructions of human metabolism is presented in Fig 4B. As illustrated in Fig 4B, HMR2 is the most comprehensive global reconstruction of human metabolism and this model together with other generic models of human metabolism has served as a basis for the reconstruction of context-specific GEMs (Fig 4C).

Context-specific GEMs were generated by manually curating the existing literature as well as by using various algorithms that have been reviewed elsewhere (Machado & Herrgard, 2014). For instance, the recently developed tINIT algorithm enables the reconstruction of simulation-ready GEMs based on proteomics data and metabolic functions that are known to occur in the cell/tissue of interest (Agren *et al*, 2014). The implementation of a metabolic function related to bile acid synthesis into the liver-specific GEM is shown as an example in Fig 4D. Recently, 32 tissue-specific GEMs for healthy human tissues were generated by integrating RNA-seq data from the HPA in combination with the tINIT algorithm and they were used to compare the metabolic differences between these

tissues (Uhlen *et al*, 2015). GEMs reconstructed based on RNA-seq data successfully predicted tissue-specific functions. For instance, the liver GEM was the only model that could successfully perform metabolic functions related to bile acid synthesis. Moreover, the liver GEM was able to perform all defined human metabolic functions and it was the largest GEM in terms of incorporated reactions, metabolites, and genes, reflecting its high metabolic activity compared to the other analyzed tissues. A list of the various cell-/tissue-specific GEMs that have been generated so far, either in physiological or in disease states, is presented in Table 3.

Applying context-dependent GEMs for analyzing human diseases

Context-specific GEMs in combination with omics data obtained in disease-specific contexts have been used to elucidate the metabolic capabilities of cells/tissues involved in metabolism-related disorders including obesity (Mardinoglu *et al*, 2013a, 2014b, 2015a), non-alcoholic fatty liver disease (NAFLD) (Mardinoglu *et al*, 2014a; Hyötyläinen *et al*, 2016), type 2 diabetes (T2D) (Varemo *et al*, 2015), and aging (Yizhak *et al*, 2013), as well as to determine unique metabolic properties of cancer cells (Agren *et al*, 2012; Gatto *et al*, 2014; Nam *et al*, 2014) and even individual cell lines (Yizhak *et al*, 2014a,b; Gatto *et al*, 2015; Ghaffari *et al*, 2015) and tumors (Agren *et al*, 2014). Each of these studies advanced our understanding of the molecular mechanisms underlying these diseases and allowed the discovery of drug targets or biomarkers that can be used for designing effective treatment strategies.

Recently, a GEM for skeletal myocytes was reconstructed using cell type-specific RNA-seq data and incorporating cell type-specific proteomics data from the HPA. First, the presence/absence of each enzyme in myocytes was determined and based on this information the corresponding metabolic reaction was incorporated into the myocyte-specific GEM (Varemo *et al*, 2015). The model was employed for characterizing the metabolic alterations in skeletal muscle in response to T2D based on the meta-analysis of six published datasets on T2D muscle gene expression. The metabolic alterations observed in the skeletal muscle T2D patients involved differences in pyruvate oxidation, tetrahydrofolate metabolism, and branched-chain amino acid catabolism.

The interplay between a large number of biological pathways and the significant variation between patients makes it extremely difficult to identify effective drug targets and biomarkers for metabolic diseases. Personalized GEMs that account for interindividual differences as well as for the unique characteristics of disease progression in each individual (Agren *et al*, 2014) present a potential solution to these issues. In a recent study, personalized cancer GEMs for six hepatocellular carcinoma (HCC) patients as well as 83 healthy cell-specific GEMs were reconstructed using the tINIT algorithm to integrate proteomics data from the HPA and metabolic functions that are present in human cells (Agren *et al*, 2014). Based on these personalized GEMs, anticancer drug targets that can be used for inhibiting the HCC tumor growth in each patient were identified. One of the targets, predicted to be effective in all patients, was experimentally validated in human HCC cancer cell lines. Overall, the observation that fat oxidation was increased in the analyzed HCC tumors indicated that targeting this

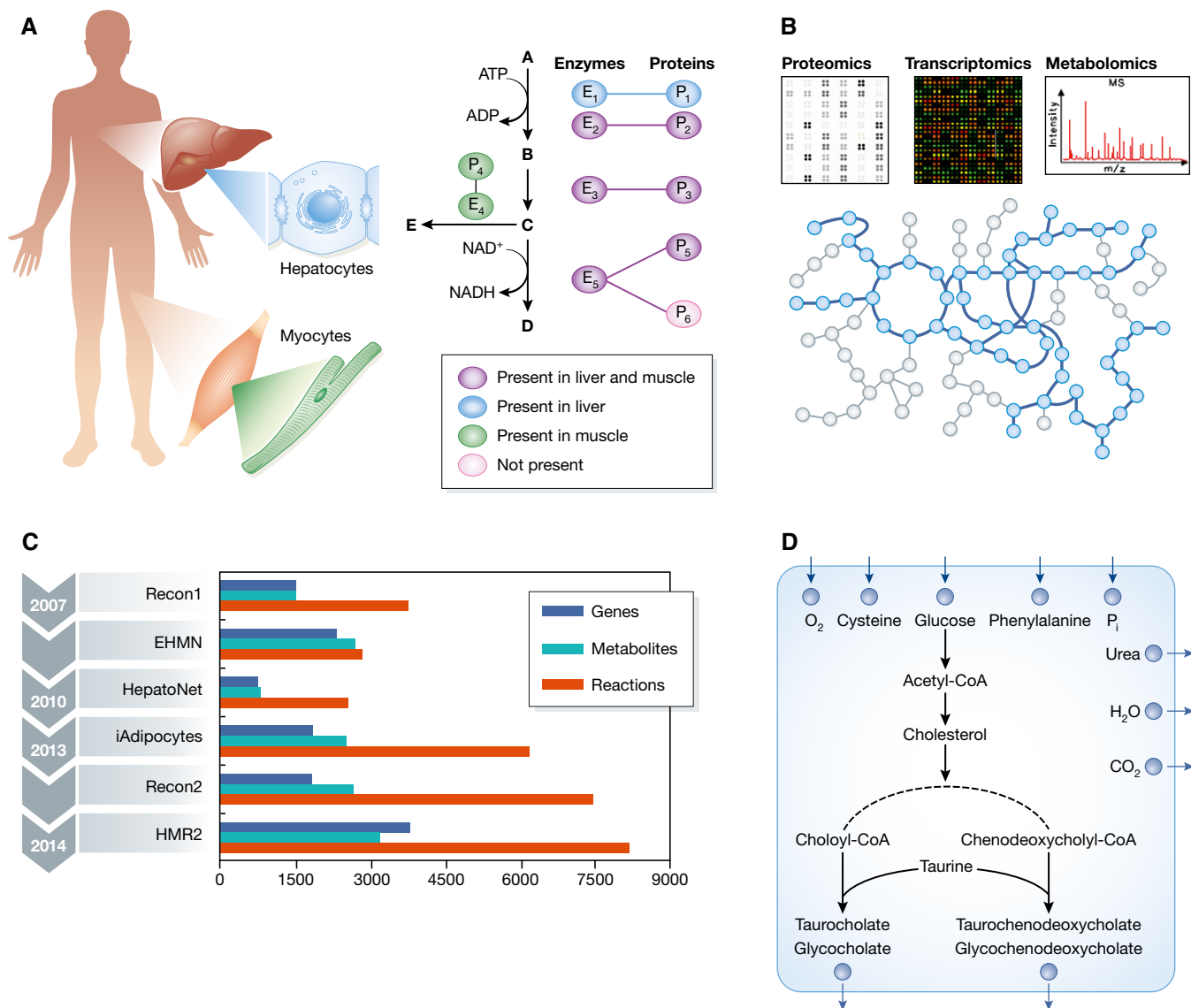


Figure 4. Genome-scale metabolic models for human cells/tissues.

(A) GEMs incorporate the known biochemical reactions and their catalyzing enzymes in a particular cell/tissue type. The information related to the reaction–gene association is used for the reconstruction of context-specific GEMs. (B) The continuously increasing number of reactions, metabolites, and genes included in generic human GEMs and manually curated cell-/tissue-specific GEMs generated in the recent years is shown. (C) High-throughput omics data including proteomics, transcriptomics, and metabolomics have been used for reconstructing cell-/tissue-specific GEMs based on generic human GEMs. (D) The metabolic tasks that are known to occur in a given human cell/tissue need to be defined to generate functional cell-/tissue-specific GEMs. The definition of the metabolic task related to bile acid synthesis in the liver is presented. Glucose, cysteine, phenylalanine, oxygen (O₂), and phosphate (P_i) must be taken up, whereas urea, water (H₂O), and carbon dioxide (CO₂) must be secreted in order to successfully simulate bile acid synthesis in liver GEM.

metabolic process could be used for developing treatment strategies for HCC.

Another recent application of GEMs in the context of HCC is presented in the study of Björnson *et al* (2015). In this case, gene expression data from approximately 360 HCC tumors and 50 non-cancerous liver samples were analyzed using a HCC-specific GEM. Interestingly, a group of patients showed an increased fat oxidation, whereas another group showed a decreased fat oxidation. The fact that HCC tumors from different patients may have completely opposite metabolic programming highlights that careful

stratification of HCC patients and personalized medicine approaches are highly advantageous for developing effective treatment strategies. Overall, these studies provide valuable insights into inter- and intratumor heterogeneity and point out that it might be extremely difficult to treat all different HCC patients with a single drug. This drug can be effective in the right context, that is, in a given patient or patient group. Therefore, personalized GEMs and their predictions of a patient's response to different drugs can be extremely useful for guiding precision medicine approaches.

Table 3. List of generic and cell-/tissue-specific human GEMs.

| Model name | Application | References |
|---|---|-----------------------------------|
| Generic human GEMs | | |
| Recon1 | Integration of genomic and bibliomic data | (Duarte <i>et al</i> , 2007) |
| EHMN | Integration of genomic and bibliomic data | (Ma <i>et al</i> , 2007) |
| HMR | Integration of previous generic human GEMs and publicly available databases | (Agren <i>et al</i> , 2012) |
| Recon2 | Community-based reconstruction of human metabolism | (Thiele <i>et al</i> , 2013) |
| HMR2 | Incorporation of extensive lipid metabolism into the generic human GEM | (Mardinoglu <i>et al</i> , 2014a) |
| Cell-/tissue-specific GEMs | | |
| Red blood cell | Analysis of the metabolic loads in red blood cells | (Wiback & Palsson, 2002) |
| Mitochondria | Study of the human mitochondrial metabolism | (Vo <i>et al</i> , 2004) |
| Fibroblasts | Metabolic alterations in Leigh syndrome | (Vo <i>et al</i> , 2007) |
| HepatoNet1 | Investigation of hepatic enzyme deficiencies | (Gille <i>et al</i> , 2010) |
| Computational liver model | Discovery of biomarkers of liver disorders including hyperammonemia and hyperglutaminemia | (Jerby <i>et al</i> , 2010) |
| Kidney | Prediction of causal drug off-targets that impact kidney function | (Chang <i>et al</i> , 2010) |
| Brain (three neuron types and astrocytes) | Revealing the metabolic alterations in Alzheimer's disease | (Lewis <i>et al</i> , 2010) |
| IAB-AMQ-1410 | Analysis of the host–pathogen interactions with <i>Mycobacterium tuberculosis</i> | (Bordbar <i>et al</i> , 2010) |
| Multitissue (hepatocytes, myocytes, and adipocytes) | Revealing the metabolic alterations in T2D | (Bordbar <i>et al</i> , 2011a) |
| Erythrocyte (iAB-RBC-283) | Revealing the complexity in the functional capabilities of human erythrocyte metabolism | (Bordbar <i>et al</i> , 2011b) |
| 69 cell-specific GEMs | Studying the metabolic differences between healthy cells and cancers | (Agren <i>et al</i> , 2012) |
| 126 tissue-specific GEMs | Comparative analysis between healthy tissues and tumor | (Wang <i>et al</i> , 2012) |
| CardioNet | The effect of oxygen and substrate supply on the efficiency of selected metabolic functions of cardiomyocytes | (Karlstaedt <i>et al</i> , 2012) |
| iAdipocytes1809 | Revealing the metabolic differences in obese subjects | (Mardinoglu <i>et al</i> , 2013a) |
| Tissue-specific GEMs | Studying the metabolic differences between healthy tissues and cancers | (Nam <i>et al</i> , 2014) |
| Liver GEM | Studying urea metabolism in liver tissue | (Vlassis <i>et al</i> , 2014) |
| 83 cell-specific GEMs | Defining the major metabolic functions in human cell types | (Agren <i>et al</i> , 2014) |
| iHepatocytes2322 | Revealing the metabolic alterations in response to NAFLD | (Mardinoglu <i>et al</i> , 2014a) |
| iMyocyte2419 | Revealing the metabolic alterations in response to T2D | (Varemo <i>et al</i> , 2015) |
| 32 tissue-specific GEMs | Global analysis of the metabolic functions in major human tissues | (Uhlen <i>et al</i> , 2015) |

Concluding remarks

Here, we reviewed some of the publicly available human transcriptomics data resources with a focus on the expression data for protein-coding genes. Tissue-restricted and tissue-enriched genes can be consistently defined in a genome-wide manner by two independent datasets generated using either fresh surgically removed tissues or postmortem tissues taken within 24 hours after the death of the individual. Thus, comprehensive lists of protein-coding genes can be compiled for all the major tissues of the human body (see Table EV1), with their quantitative expression profiles generated by deep sequencing of the transcriptome.

The use of high-quality proteomics and transcriptomics data in combination with metabolic modeling allows for functional analyses in the context of different pathologies, for example, by comparing GEMs reconstructed using data from healthy and diseased subjects. On the one hand, a comparison of the healthy vs. diseased GEM

topology can provide insights into how cancer metabolism differs from metabolism of the healthy tissue (Gatto *et al*, 2014; Björnson *et al*, 2015; Zhang *et al*, 2015). Furthermore, GEMs can be used for identifying drug targets (Agren *et al*, 2014), and therefore, their integration with omics data generated in a clinical setup can be applied to guide precision medicine in different disease types. Further improvement and expansion of GEMs to cover other biological processes, for example, protein secretion pathways and protein synthesis (Feizi *et al*, 2013), will allow this modeling framework to capture dysfunction of key cellular pathways in a range of different pathologies, potentially leading to the identification of new treatment strategies and biomarkers.

The transcriptomics data can be complemented with immunohistochemistry to define protein localization in the subcompartments of each tissue and organ down to the single-cell level (www.proteinatlas.org). Moreover, extending these tissue profiles to include splice variants and protein modifications is important for improving

our understanding of the role of the isoform proteome and post-translational modifications in human physiology and disease. Finally, spatial proteomics using fluorescent-based antibody profiling (Marx, 2015) can provide even higher resolution with precise localizations of the corresponding proteins down to subcellular compartments and various substructures. The integration of transcriptomics data with other large-scale data, such as mass spectrometry-based proteomics, antibody-based profiling, and metabolomics, can thus generate an important molecular knowledge base for systems biology of human health and disease.

Expanded View for this article is available online.

Acknowledgements

We acknowledge the entire staff of the Human Protein Atlas program and the Science for Life Laboratory for valuable contributions. Funding was provided by the Knut and Alice Wallenberg Foundation.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J (2012) Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput Biol* 8: e1002518
- Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J (2014) Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol Syst Biol* 10: 721
- Anderson L, Seilhamer J (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18: 533–537
- Bahcall OG (2015) Human genetics: GTEx pilot quantifies eQTL variation across tissues and individuals. *Nat Rev Genet* 16: 375
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 41: D991–D995
- Bjornson E, Boren J, Mardinoglu A (2016) Personalized Cardiovascular Disease Prediction and Treatment—A Review of Existing Strategies and Novel Systems Medicine Tools. *Front Physiol* 7: 2
- Björnson E, Mukhopadhyay B, Asplund A, Pristovsek N, Cinar R, Romeo S, Uhlen M, Kunos G, Nielsen J, Mardinoglu A (2015) Stratification of hepatocellular carcinoma patients based on acetate utilization. *Cell Rep* 13: 2014–2026
- Bordbar A, Lewis NE, Schellenberger J, Palsson BO, Jamshidi N (2010) Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Mol Syst Biol* 6: 422
- Bordbar A, Feist AM, Usaite-Black R, Woodcock J, Palsson BO, Famili I (2011a) A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. *BMC Syst Biol* 5: 180
- Bordbar A, Jamshidi N, Palsson BO (2011b) iAB-RBC-283: a proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states. *BMC Syst Biol* 5: 110
- Bossi A, Lehner B (2009) Tissue specificity and the human protein interaction network. *Mol Syst Biol* 5: 260
- Brawand D, Soumilion M, Necsulea A, Julien P, Csardi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW, Zeller U, Khaitovich P, Grutzner F, Bergmann S, Nielsen R, Paabo S, Kaessmann H (2011) The evolution of gene expression levels in mammalian organs. *Nature* 478: 343–348
- Chang RL, Xie L, Bourne PE, Palsson BO (2010) Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput Biol* 6: e1000938
- GTEx Consortium (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348: 648–660
- Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Giron CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Kahari AK, Keenan S *et al* (2015) Ensembl 2015. *Nucleic Acids Res* 43: D662–D669
- Danielsson F, James T, Gomez-Cabrero D, Huss M (2015) Assessing the consistency of public human tissue RNA-seq data sets. *Brief Bioinform* 16: 941–949
- Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* 104: 1777–1782
- Eden E, Geva-Zatorsky N, Issaeva I, Cohen A, Dekel E, Danon T, Cohen L, Mayo A, Alon U (2011) Proteome half-life dynamics in living human cells. *Science* 331: 764–768
- Edqvist PH, Fagerberg L, Hallstrom BM, Danielsson A, Edlund K, Uhlen M, Ponten F (2015) Expression of human skin-specific genes defined by transcriptomics and antibody-based profiling. *J Histochem Cytochem* 63: 129–141
- Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpour S, Danielsson A, Edlund K, Asplund A, Sjostedt E, Lundberg E, Szigartyo CA, Skogs M, Takanen JO, Berling H, Tegel H, Mulder J, Nilsson P *et al* (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* 13: 397–406
- Feizi A, Osterlund T, Petranovic D, Bordel S, Nielsen J (2013) Genome-scale modeling of the protein secretory machinery in yeast. *PLoS ONE* 8: e63284
- Frezza C, Zheng L, Folger O, Rajagopalan KN, MacKenzie ED, Jerby L, Micaroni M, Chaneton B, Adam J, Hedley A, Kalna G, Tomlinson IP, Pollard PJ, Watson DG, DeBerardinis RJ, Shlomi T, Ruppin E, Gottlieb E (2011) Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature* 477: 225–228
- Gatto F, Nookaew I, Nielsen J (2014) Chromosome 3p loss of heterozygosity is associated with a unique metabolic network in clear cell renal carcinoma. *Proc Natl Acad Sci USA* 111: E866–E875
- Gatto F, Miess H, Schulze A, Nielsen J (2015) Flux balance analysis predicts essential genes in clear cell renal cell carcinoma metabolism. *Sci Rep* 5: 10738
- Ghaffari P, Mardinoglu A, Asplund A, Shoaie S, Kampf C, Uhlen M, Nielsen J (2015) Identifying anti-growth factors for human cancer cell lines through genome-scale metabolic modeling. *Sci Rep* 5: 8183
- Gibson G (2015) Human genetics. GTEx detects genetic effects. *Science* 348: 640–641
- Gille C, Bolling C, Hoppe A, Bulik S, Hoffmann S, Hubner K, Karlstadt A, Ganeshan R, Konig M, Rother K, Weidlich M, Behre J, Holzthutter HG (2010) HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol Syst Biol* 6: 411
- Gry M, Rimini R, Stromberg S, Asplund A, Ponten F, Uhlen M, Nilsson P (2009) Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genom* 10: 365

- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C et al (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22: 1760–1774
- Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, van de Lagemaat LN, Smith KA, Ebbert A, Riley ZL, Abajian C, Beckmann CF, Bernard A, Bertagnolli D, Boe AF, Cartagena PM, Chakravarty MM, Chapin M, Chong J, Dalley RA et al (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489: 391–399
- Hebenstreit D, Fang M, Gu M, Charoensawan V, van Oudenaarden A, Teichmann SA (2011) RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol Syst Biol* 7: 497
- Hyötyläinen T, Jerby L, Petäjä EM, Mattila I, Jäntti S, Auvinen P, Gastaldelli A, Yki-Järvinen H, Ruppén E, Orešić M (2016) Genome-scale study reveals reduced metabolic adaptability in patients with non-alcoholic fatty liver disease. *Nat Commun* 7: 8994
- Jerby L, Shlomi T, Ruppén E (2010) Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol Syst Biol* 6: 401
- Kampf C, Mardinoglu A, Fagerberg L, Hallström B, Edlund K, Nielsen J, Uhlen M (2014) The human liver-specific proteome defined by transcriptomics and antibody-based profiling. *FASEB J* 28: 2901–2914
- Karlstaedt A, Fliegner D, Kararigas G, Ruderisch HS, Regitz-Zagrosek V, Holzhtuter HG (2012) CardioNet: a human metabolic network suited for the study of cardiomyocyte metabolism. *BMC Syst Biol* 6: 114
- Keen JC, Moore HM (2015) The Genotype-Tissue Expression (GTEx) Project: Linking Clinical Data with Molecular Analysis to Advance Personalized Medicine. *J Pers Med* 5: 22–29
- Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabudhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LD et al (2014) A draft map of the human proteome. *Nature* 509: 575–581
- Krupp M, Marquardt JU, Sahin U, Galle PR, Castle J, Teufel A (2012) RNA-Seq Atlas—a reference database for gene expression profiling in normal tissue by next-generation sequencing. *Bioinformatics* 28: 1184–1185
- Lamond AI, Uhlen M, Horning S, Makarov A, Robinson CV, Serrano L, Hartl FU, Baumeister W, Werenskiöld AK, Andersen JS, Vorm O, Linnal M, Aebersold R, Mann M (2012) Advancing cell biology through proteomics in space and time (PROSPECTS). *Mol Cell Proteomics* 11: O112. O17731
- Lewis NE, Schramm G, Bordbar A, Schellenberger J, Andersen MP, Cheng JK, Patel N, Yee A, Lewis RA, Eils R, König R, Palsson BO (2010) Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nat Biotechnol* 28: 1279–1285
- Lundberg E, Fagerberg L, Klevebring D, Matic I, Geiger T, Cox J, Algenas C, Lundberg J, Mann M, Uhlen M (2010) Defining the transcriptome and proteome in three functionally different human cell lines. *Mol Syst Biol* 6: 450
- Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I (2007) The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol* 3: 135
- Machado D, Herrgard M (2014) Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput Biol* 10: e1003580
- Maier T, Guell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Lett* 583: 3966–3973
- Maier T, Schmidt A, Guell M, Kuhner S, Gavin AC, Aebersold R, Serrano L (2011) Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol Syst Biol* 7: 511
- Mardinoglu A, Nielsen J (2012) Systems medicine and metabolic modelling. *J Intern Med* 271: 142–154
- Mardinoglu A, Agren R, Kampf C, Asplund A, Nookaew I, Jacobson P, Walley AJ, Froguel P, Carlsson LM, Uhlen M, Nielsen J (2013a) Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol Syst Biol* 9: 649
- Mardinoglu A, Gatto F, Nielsen J (2013b) Genome-scale modeling of human metabolism – a systems biology approach. *Biotechnol J* 8: 985–996
- Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J (2014a) Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat Commun* 5: 3083
- Mardinoglu A, Kampf C, Asplund A, Fagerberg L, Hallström BM, Edlund K, Blüher M, Ponten F, Uhlen M, Nielsen J (2014b) Defining the human adipose tissue proteome to reveal metabolic alterations in obesity. *J Proteome Res* 13: 5106–5119
- Mardinoglu A, Nielsen J (2015) New paradigms for metabolic modeling of human cells. *Curr Opin Biotech* 34: 91–97
- Mardinoglu A, Heiker JT, Gärtner D, Björnson E, Schön MR, Flehmig G, Klötting N, Krohn K, Fasshauer M, Stummvoll M, Nielsen J, Blüher M (2015a) Extensive weight loss reveals distinct gene expression changes in human subcutaneous and visceral adipose tissue. *Sci Rep* 5: 14841
- Mardinoglu A, Shoaie S, Bergentall M, Ghaffari P, Zhang C, Larsson E, Bäckhed F, Nielsen J (2015b) The gut microbiota modulates host amino acid and glutathione metabolism in mice. *Mol Syst Biol* 11: 834
- Marx V (2015) Mapping proteins with spatial proteomics. *Nat Methods* 12: 815–819
- Nam H, Campodonico M, Bordbar A, Hyduke DR, Kim S, Zielinski DC, Palsson BO (2014) A systems approach to predict oncometabolites via context-specific genome-scale metabolic networks. *PLoS Comput Biol* 10: e1003837
- Nilsson CL, Mostovenko E, Lichti CF, Ruggles K, Fenyo D, Rosenbloom KR, Hancock WS, Paik YK, Omenn GS, LaBaer J, Kroes RA, Uhlen M, Hober S, Vegvari A, Andren PE, Sulman EP, Lang FF, Fuentes M, Carlsson E, Emmett MR et al (2015) Use of ENCODE resources to characterize novel proteoforms and missing proteins in the human proteome. *J Proteome Res* 14: 603–608
- O'Brien EJ, Monk JM, Palsson BO (2015) Using Genome-scale Models to Predict Biological Capabilities. *Cell* 161: 971–987
- Papin JA, Hunter T, Palsson BO, Subramaniam S (2005) Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol* 6: 99–111
- Petryszak R, Keays M, Tang YA, Fonseca NA, Barrera E, Burdett T, Fullgrave A, Fuentes AM, Jupp S, Koskinen S, Mannion O, Huerta L, Megy K, Snow C, Williams E, Barzine M, Hastings E, Weisser H, Wright J, Jaiswal P et al (2015) Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res* 44: D746–D752
- Qian J, Esumi N, Chen Y, Wang Q, Chowdhury I, Zack DJ (2005) Identification of regulatory targets of tissue-specific transcription factors: application to retina-specific gene regulation. *Nucleic Acids Res* 33: 3479–3491
- Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Ison J, Keays M, Kurbatova N, Malone J, Mani R, Mupo A, Pedro Pereira R, Pilicheva E, Rung J, Sharma A, Tang YA, Terment T et al (2013) ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res* 41: D987–D990

- Schwanhaussier B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. *Nature* 473: 337–342
- Stadler C, Rexhepaj E, Singan VR, Murphy RF, Pepperkok R, Uhlen M, Simpson JC, Lundberg E (2013) Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. *Nat Methods* 10: 315–323
- Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6: 377–382
- Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, Thorleifsson SG, Agren R, Bolling C, Bordel S, Chavali AK, Dobson P, Dunn WB, Endler L, Hala D, Hucka M et al (2013) A community-driven global reconstruction of human metabolism. *Nat Biotechnol* 31: 419–425
- Tian Q, Stepaniants SB, Mao M, Weng L, Feetham MC, Doyle MJ, Yi EC, Dai H, Thorsson V, Eng J, Goodlett D, Berger JP, Gunter B, Linseley PS, Stoughton RB, Aebersold R, Collins SJ, Hanlon WA, Hood LE (2004) Integrated genomic and proteomic analyses of gene expression in Mammalian cells. *Mol Cell Proteomics* 3: 960–969
- Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigartyo CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T et al (2015) Proteomics. Tissue-based map of the human proteome. *Science* 347: 1260419
- UniProt C (2015) UniProt: a hub for protein information. *Nucleic Acids Res* 43: D204–D212
- Varemo L, Scheele C, Broholm C, Mardinoglu A, Kampf C, Asplund A, Nookaew I, Uhlen M, Pedersen BK, Nielsen J (2015) Proteome- and transcriptome-driven reconstruction of the human myocyte metabolic network and its use for identification of markers for diabetes. *Cell Rep* 11: 921–933
- Vlassis N, Pacheco MP, Sauter T (2014) Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput Biol* 10: e1003424
- Vo TD, Greenberg HJ, Palsson BO (2004) Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data. *J Biol Chem* 279: 39532–39540
- Vo TD, Paul Lee WN, Palsson BO (2007) Systems analysis of energy metabolism elucidates the affected respiratory chain complex in Leigh's syndrome. *Mol Genet Metab* 91: 15–22
- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470–476
- Wang YL, Eddy JA, Price ND (2012) Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst Biol* 6: 153
- Wiback SJ, Palsson BO (2002) Extreme pathway analysis of human red blood cell metabolism. *Biophys J* 83: 808–818
- Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, Reimer U, Wenschuh H, Mollenhauer M, Slotta-Huspenina J, Boese JH, Bantscheff M, Gerstmair A et al (2014) Mass-spectrometry-based draft of the human proteome. *Nature* 509: 582–587
- Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, Hodge CL, Haase J, Janes J, Huss JW 3rd, Su AI (2009) BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* 10: R130
- Yizhak K, Gabay O, Cohen H, Ruppin E (2013) Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nat Commun* 4: 2632
- Yizhak K, Gaude E, Le Devedec S, Waldman YY, Stein GY, van de Water B, Frezza C, Ruppin E (2014a) Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. *eLife* 3: e03641
- Yizhak K, Le Devedec SE, Rogkoti VM, Baenke F, de Boer VC, Frezza C, Schulze A, van de Water B, Ruppin E (2014b) A computational study of the Warburg effect identifies metabolic targets inhibiting cancer migration. *Mol Syst Biol* 10: 744
- Yizhak K, Chaneton B, Gottlieb E, Ruppin E (2015) Modeling cancer metabolism on a genome scale. *Mol Syst Biol* 11: 817
- Yu NY, Hallstrom BM, Fagerberg L, Ponten F, Kawaji H, Carninci P, Forrest AR, Fantom CT, Hayashizaki Y, Uhlen M, Daub CO (2015) Complementing tissue characterization by integrating transcriptome profiling from the Human Protein Atlas and from the FANTOM5 consortium. *Nucleic Acids Res* 43: 6787–6798
- Zhang C, Ji B, Mardinoglu A, Nielsen J, Hua Q (2015) Logical transformation of genome-scale metabolic models for gene level applications and analysis. *Bioinformatics* 31: 2324–2331



License: This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.